MULTIMEDIA UNIVERSITY

STUDENT ID NO.

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, 2022/2023

BMD3024 – MULTIVARIATE DATA ANALYSIS

(All sections / Groups)

12 JULY 2023
9.00 a.m. - 11.00 a.m.
(2 Hours)

INSTRUCTIONS TO STUDENTS

1. This question paper consists of **ELEVEN (11)** printed pages with **FOUR (4)** structured questions including the cover page.
2. Marks are shown at the end of each question.
3. Write all your answers in the answer booklet provided.
4. Answer **ALL** questions.

Question One (25 marks)

(a) Assess the **FOUR (4)** sources of bias and common problems that may be faced in conducting logistic regression and how these issues could be solved.

(16 marks)

(b) A group of researchers investigated whether household chaos predicted children's problem behaviour over and above parenting. Data was collected from 118 families. A linear regression was conducted with the variables measuring mother–child positivity (*pos*), mother–child negativity (*neg*), parent–child warmth (*warmth*), parent–child anger (*anger*); and household chaos (*chaos*) as independent variables and child's adjustment (*sdq*) as the dependent variable.

Table 1: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.316[a]	.100	.050	.17051	2.025

(i) Interpret **Table 1: Model Summary**.

(2 marks)

Table 2: ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	**Regression**	.290	5	.058	1.996	.087[b]
	Residual	2.617	90	.029		
	Total	2.907	95			

a. Dependent Variable: sdq

b. Predictors: (Constant), chaos, pos, anger, neg, warmth

(ii) Assess **Table 2: ANOVA**. Is the *F*-statistic significant at 5 percent level? What does this indicate?

(4 marks)

Continued...

Table 3: Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
x	(Constant)	.877	.288		3.050	.003	.306	1.449		
	pos	.002	.005	.048	.463	.644	-.008	.013	.949	1.053
	neg	.011	.006	.194	1.820	.072	-.001	.023	.884	1.131
	warmth	-.001	.024	-.005	-.044	.965	-.049	.047	.855	1.170
	anger	.002	.017	.009	.091	.928	-.032	.035	.939	1.064
	chaos	.074	.036	.215	2.072	.041	.003	.144	.933	1.072

a. Dependent Variable: sdq

(iii) Assess **Table 3: Coefficients.** Which are the variables with worse multicollinearity scores than the others?

(3 marks)

Continued...

Question Two (25 marks)

A researcher is interested in how three independent variables, namely Graduate Record Exam (GRE) scores (*gre*) which is required to enter graduate school (postgraduate studies) in the United States, undergraduate Cumulative Grade Point Average (CGPA) (*gpa*) and prestige of the undergraduate institution (*rank*), affect success of admission into graduate school (*admit*). The dependent variable *admit* is a binary variable (Admit vs. Don't Admit).

Table 4: Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	400	100.0
	Missing Cases	0	.0
	Total	400	100.0
Unselected Cases		0	.0
Total		400	100.0

a. If weight is in effect, see classification table for the total number of cases.

(a) Assess **Table 4: Case Processing Summary** to assess if there are any missing data. (2 marks)

Continued...

Table 5: Dependent Variable Encoding

Original Value	Internal Value
Do Not Admit	0
Admit	1

Table 6: Classification Table[a,b]

	Observed		Predicted		
			admit		Percentage Correct
			Do Not Admit	Admit	
Step 0	admit	Do Not Admit	273	0	100.0
		Admit	127	0	.0
	Overall Percentage				68.3

a. Constant is included in the model.

b. The cut value is .500

(b) Referring to **Table 5: Dependent Variable Encoding** and **Table 6: Classification Table**, interpret the coding of the two categories of success of admission into graduate school, the number of students in each of the two categories, and the overall percentage when the **base model (Step 0)** is applied.

(4 marks)

Table 7: Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.765	.107	50.764	1	<.001	.465

(c) Referring to **Table 7: Variables in the Equation**, assess if the base model is able to predict make predictions that are significantly different than by chance (50-50).

(2 marks)

Continued...

Table 8: Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	40.535	3	<.001
	Block	40.535	3	<.001
	Model	40.535	3	<.001

(d) Referring to **Table 8: Omnibus Tests of Model Coefficients**, interpret if the logistic regression model that includes the three independent variables is significant.

(4 marks)

Table 9: Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	459.442[a]	.096	.135

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

(e) Interpret **Table 9: Model Summary.** (4 marks)

Table 10: Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	3.249	8	.918

(f) Based on **Table 10: Hosmer and Lemeshow Test**, assess whether the null hypothesis 'Predictions made by the model fit well with observed group memberships' is rejected.

(2 marks)

Continued...

Table 11: Classification Table[a]

	Observed		Predicted		
			admit		Percentage Correct
			Do Not Admit	Admit	
Step 1	admit	Do Not Admit	253	20	92.7
		Admit	98	29	22.8
	Overall Percentage				70.5

a. The cut value is .500

(g) According to **Table 11: Classification Table**, interpret the number of students predicted in each of the two categories, and the overall percentage when the **model with three predictors (Step 1)** is applied.

(2 marks)

Table 12: Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	gre	.002	.001	4.414	1	.036	1.002	1.000	1.004
	gpa	.777	.327	5.630	1	.018	2.175	1.145	4.132
	rank	-.560	.127	19.404	1	<.001	.571	.445	.733
	Constant	-3.450	1.133	9.272	1	.002	.032		

a. Variable(s) entered on step 1: gre, gpa, rank.

(h) Assess from **Table 12: Variables in the Equation** if any of the three independent variables are significant in determining whether the respondent gets admitted into graduate school.

(2 marks)

(i) Based on **Table 12**, compute the Percentage Change in Odds of predicting correctly whether a certain student will be admitted into graduate school, if GRE score (*gre*) is known.

(3 marks)

Continued...

Question Three (25 marks)

(a) A researcher conducted a Principal Component Analysis on twenty-eight variables. The following tables are based on analysis of the all those variables.

Table 13: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.894
Bartlett's Test of Sphericity	**Approx. Chi-Square**	2989.769
	df	378
	Sig.	.000

(i) Based on **Table 13: KMO and Bartlett's Test**, assess whether factor analysis can be carried out.

(5 marks)

Continued...

Table 14: Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
q1	.409	.545		.351	
q2				.765	
q3		.666	.355		
q4		.619	.306	.321	
q5					.666
q6				.781	
q7		.399	.535		
q8	.587	.383	.362		
q9	.432	.470		.447	
q10	.440			.450	
q11	.386	.345	.435		
q12					.506
q13	.593	.371			
q14	.764				
q15	.447	.509			
q16	.775				
q17	.714				
q18				.586	.316
q19			.712		
q20			.649		
q21	.505	.523			
q22	.748	.338			
q23					.672
q24		.744			
q25			.688		
q26	.407		.568		
q27			.613		.411
q28					.673

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 7 iterations.

(ii) Referring to **Table 14: Rotated Component Matrix**, produce a list all the variables for each of the **FIVE (5)** factors (i.e. **Component 1** to **Component 5**), based on their factor scores.

(13 marks)

Continued...

(b) A Student Affairs officer would like to investigate the number of underclassmen (juniors) and upperclassmen (seniors) (variable name: *Rank_2*) that live on campus compared to off campus (variable name: *LiveOnCampus*). She surveyed 388 students and obtained the following summary upon conducting Chi-Square analysis.

Table 15: Rank_2 * LiveOnCampus Crosstabulation

			LiveOnCampus		Total
			Off-campus	On-campus	
Rank_2	Underclassmen	Count	79	148	227
		Expected Count	135.1	91.9	227.0
		% within Rank_2	34.8%	65.2%	100.0%
		% within LiveOnCampus	34.2%	94.3%	58.5%
		% of Total	20.4%	38.1%	58.5%
	Upperclassmen	Count	152	9	161
		Expected Count	95.9	65.1	161.0
		% within Rank_2	94.4%	5.6%	100.0%
		% within LiveOnCampus	65.8%	5.7%	41.5%
		% of Total	39.2%	2.3%	41.5%
Total		Count	231	157	388
		Expected Count	231.0	157.0	388.0
		% within Rank_2	59.5%	40.5%	100.0%
		% within LiveOnCampus	100.0%	100.0%	100.0%
		% of Total	59.5%	40.5%	100.0%

Compute the following odds based on **Table 15: Rank_2 * LiveOnCampus Crosstabulation**:

(i) A student lives on campus if he is an underclassman.

(2 marks)

(ii) A student lives on campus if he is an upperclassman.

(2 marks)

(iii) The Odds Ratio of a student lives on campus if they are an underclassman versus an upperclassman.

(3 marks)

Continued...

Question Four (25 Marks)

(a) A researcher conducted Principal Component Analysis on twenty variables. Four variables, namely, *v10*, *v15*, *v18*, and *v20* were removed after analysing the correlation matrix. **Table 16** shows the independent variables that belong to Factor 1 (*F1*) to Factor 4 (*F4*) independent factors:

Table 16

Factor 1 (F1)	Factor 2 (F2)	Factor 3 (F3)	Factor 4 (F4)
V2	V6	V1	V4
V9	V7	V3	V12
V13	V8	V5	V14
V16	V11	V20	--
V17	--	--	--

Sketch **ONE (1)** path diagram that shows the structural relationship with the independent factors above with the Overall Customer Satisfaction (*overall*) as the dependent variable.

(13 marks)

(b) Interpret **ALL** the model fit statistics found in **Figure 1**. Is this a good model?

(12 marks)

Figure 1

ChiSquare=127.221
DF=116
ChiSquareDF=1.097
p=.224
Ratio=.758
RMSEA=.016
TLI=.988
CFI=.991

End of Page.